Linked Open Data を活用した電子書籍に関する研究

A Study on E-books utilizing Linked Open Data

宮田惇平
Junpei MIYATA
指導教員　藤井章博

法政大学大学院工学研究科情報電子工学専攻修士課程

With popularization of tablet-type device and E-books, a new style of editing books is emerging. This makes it easier for us to edit and publish e-books and modify their contents. Currently, majority of E-books are not much more than transform of paper books. However, there is a high potential in utilizing open data, that is going to be provided through used in wide range of fields. This paper focuses on utilizing “Linked Open Data” for XHTML used by EPUB in terms of using open data. In this report, we propose the editing E-books application which is interactive and contains contents on the clearly relation of data.

Key Words *:E-books,electric publishing, Linked Open Data, XHTML, Authoring System*

１．はじめに

近年，電子書籍が登場し，書籍，取扱説明書等がデジタル化されるとともに，ネットワークを通じた流通にも大きな変化が生じた．電子書籍の普及に伴い，タブレット端末やインターネットを利用した電子書籍の運用も重要視される．また，教育の場においてもタブレット・電子書籍を導入する学校も登場した.電子書籍市場は現状，紙を単に電子書籍化したものの売上が大多数である．一方で，紙の書籍には無い利便性を持つ電子書籍（本稿ではデジタルオンリーと呼ぶ）も登場し，今後デジタルオンリーな電子書籍の製作が盛んになると考えられる．

データの公開に関しては例えば，2013 年の G8 サミットで「オープンデータ憲章」[1]が制定され，オープンデータは国際的にも注目を浴びつつあり，世界規模でその貯蓄と利活用が進められている．また，福井県鯖江市[2]といった地方自治体でもオープンデータの利活用を促進している．本研究では，公開されているデータを活用するという観点から，Linked Open Data[3](LOD)に着目した．LOD を活用する研究は様々あるが，電子書籍に活用する研究については，体系に関するもの[4]や，読者側に関する研究[5,6]があり，製作・編集に関しては研究の余地があると考える．そこで，本研究では電子書籍の形式の一つである EPUB 形式[7]に LOD を活用したデジタルオンリーなコンテンツの編集に関しての研究を行った．

2 では電子書籍におけるオープンデータの活用について述べる．3 では関連技術について述べる． 4 では本稿で示すオーサリングシステムの設計・実装について述べる．5 では，一例として LOD を公開している DBpedia Japanese[8]と日本語Wikipediaオントロジー[9]を利用し，日本の 100 名城のコンテンツ製作と閲覧を行うデジタルオンリーで，ハイパーテキストとしての特性を活かしたアプリケーションの開発を行った．6 では前述のアプリケーションに関して考察を行う．7 はむすびとする．

２．電子書籍におけるオープンデータの活用

（１）書籍の制作

新しい書籍の形態と読書の方法を議論するために，書籍の一般的な製作過程を「執筆」「編集」「出版」の 3 つの工程に分け，製作に関わる者を「著者」「読者」の 2 つの主体に着目し検討を行った．まず、図 1 に従来の書籍の製作過程を整理する．次に電子書籍によって製作過程に変化が生まれると考え、その変化の可能性について検討する．

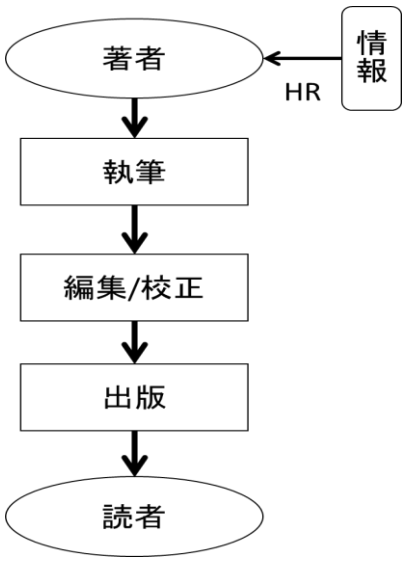

図 1.　従来の製作過程

従来の過程では，製作過程はソフトウェア開発における，ウォーターフォールモデルに近いと考えられる．執筆には HR データを使う．著者と読者の関係性は非常に遠く、読者は出版された本を購入、借入して読むだけである．図 2 に本稿で想定する新しい書籍の製作過程を示す。

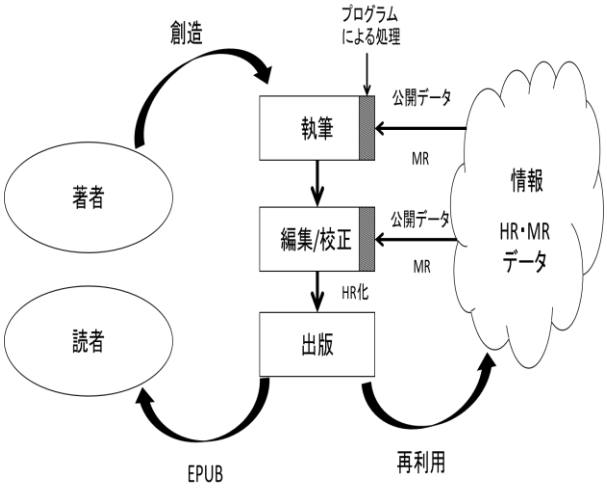

図 2. 電子書籍の新しい製作過程

以下それぞれの項目を説明する．

- 執筆

執筆活動に HR・MR データを利用する事が可能となる．この点が特に本稿で注目する観点である．

- 編集／校正

編集工程において、著者の創造性に基づく編集機能を盛り込み、MR データをプログラムによる処理を行うことで、出版物を構成する機能を具備することが可能である。

- 出版

出版された電子書籍は読むだけでなく、公開データとして再利用する事が可能である。

- 著者

必要なデータを Web 上から取得し、プログラムによる処理を行う事で、執筆・編集の自動化が可能となり、著者の負担は軽くなると考えられる．

- 読者

従来は、メディアを通じた情報の公開とは既存の書籍において、ほぼ一方向であると言える。一方、新過程では、公開データを執筆に利用することで、SNS 等 Web 上での情報共有の場が広がり、書物を中心とする読者と著者との間の強い関係性が生じる事も期待できる。

本稿では執筆活動に LOD を活用する事を想定した。LOD は MR データである。著者の視点から、執筆・編集工程における、MR データのプログラム処理を自動化できれば、負担も軽くなり、また作業の効率化にも繋がる。そこで、本稿では新しい製作体系に基づく電子書籍制作支援アプリケーションを実装した。アプリケーション設計に重要となる技術について 3 で示す．

3．関連技術

（1）EPUB

EPUB とは，電子書籍の規格の 1 つである．米国の電子書籍の標準化団体の 1 つである International Digital Publishing Forum (IDPF) が提唱した公開された仕様の電子書籍用ファイル・フォーマット規格である．EPUB のコンテンツ には Extensible HyperText Markup Language(XHTML)形式が採択されている。XHTML は HTML を XML の記述形式に添って記述した形式である．このため構成化した文書として Document Object Model(DOM)によって形成される部分の集合体として書籍が制作できる．このことは，言語中立であり、コンテンツの特定の部分を簡単に抽出・変更・削除が可能な製作過程を実現できる事を意味する．また，XHTML では、図 3 に示すように，HTML だけでなく，画像や音声，動画と言ったマルチメディアを利用する事が可能である．マルチメディア利用することで紙の本と差別化を図る電子書籍も存在する．

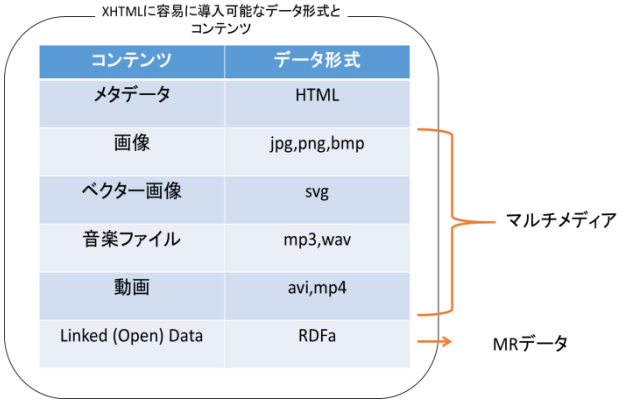

図 3.　XHTML に導入可能なコンテンツとその形式

オープンデータ活用が進み，データとして Human Readable(HR)・Machine Readable(MR)なデータが増大すると考えられる．また，コンテンツにも HR データだけでなく，MR を活用する機会が増えてくる．そこで，デジタルオンリーなコンテンツを開発する場合，資源となるデータの特性を意識する必要がある．そこで以下では Web で提供されるデータを HR・MR の 2 つの観点で分類し，電子書籍の素材とすることを検討した．

（２）Human Readable・Machine Readable

人間が読解することを前提として公開されている情報を Human Readable(HR)データと呼ぶ．Web はこの HR データの集合であり，今でもその情報は増え続けている．これは，情報爆発と呼ばれるほどであり，ユーザが目視のみで情報を精査し必要なものを抽出するのは困難である．これに対し，プログラムの処理に対応できるデータとして Machine Readable(MR)データがある．ここで，HR データに対し MR データであるセマンティック情報を付加することで Web の利便性を高めることを目的としたものがセマンティック Web といえる．セマンティック Web は W3C によって提唱され，RDF（Resource Description Framework）や OWL（Web Ontology Language）のようなセマンティクスを扱うツールを用いて Web ページに意味情報を持たせている．

（３）セマンティック Web

セマンティック Web[10]とは，W3C が提唱する次世代のウェブの形である．Web ページおよびその中に記述された内容について，情報(メタデータ)を規則に従って付加することで，コンピュータが効率よく情報を収集・解釈できるようにする体系である．インターネットをデータの集合から知識のデータベースに進化させようという試みといえる．以下の Linked Open Data(LOD)はその一例で，現在世界規模でその蓄積と利活用が進んでいる．

（４）LOD

LOD は，Web の技術を利用して，計算機が処理しやすい形式(MR)で情報を共有する仕組みである．インターネット上のオープンな場へ LOD の形式で発信することで，情報を多くの人々へ広くかつ迅速に伝えることが可能となる．また，発信された情報を，Web 上で共有し，リンクすることによって，Web 上に巨大な知識データベースが形成されている．こうした知識を利用することで，価値ある新しいサービスの可能性がある．例えば，国内でも，DBpedia Japanese，日本語 Wikipedia オントロジー，LODAC[11]など LOD としてデータセットを公開しているプロジェクトが存在する．現在，LOD の為のデータセットや，LOD を活用したアプリケーションが加速度的に増加している．本研究ではこの LOD を電子書籍のコンテンツに利用する．

（５）RDF・RDFa

LOD を支える要素技術が Resource Description Framework(RDF)である。RDF は、Web 上でリソースに関する情報を表すための言語で，リソースの関係を主語、述語、目的語という 3 つの要素（トリプル）で表現する．RDF は、人間に表示するだけではなく、アプリケーションが情報を処理する必要のある状況を目的としている．

RDF は、この情報を表現するための共通の枠組みを提供するため、意味を損なわずにアプリケーション間で情報交換が可能である．共通の枠組みであるため、アプリケーションの設計者は共通の RDF パーサや処理ツールを有効利用でき，異なるアプリケーション間で情報交換できるということは、情報が元々作成された以外のアプリケーションでその情報を利用できることを意味する．

W3C によって簡便なセマンティックの表現形式として規定されているものの一つに RDFa がある．RDFa は RDF のシリアライゼーションフォーマットの 1 つである．これは，XHTML のタグ中の要素としてメタデータを埋め込むものであり，これを利用することで Web ページは HR データでありながら MR データとしても容易に扱えるようになる．また，RDFa では XHTML の仕様を柔軟に活用することができる．これによって，セマンティックを表現した MR リソースに対応するページ，本稿では LOD を公開しているサイトを HR コンテンツとして表示することができる． RDFa は，HR・MR データであるためどちらの機能でも出力として活用できる．

次の 4 では，LOD を EPUB に組み込む手法，アプリケーションの設計について述べていく．

４．オーサリングシステムの設計・実装

（１）設計

2 で述べた書籍製作の新しい製作過程のモデルに，3 で示した関連技術を反映させる Web アプリケーションの設計を述べる．

- 執筆
  外部リソースから読者や編集者のニーズの反映に有効なリソースを柔軟かつ容易に獲得する必要がある．本稿では対象の外部リソースを公開データ(MR)とする．
- 編集/校正
  入稿で得られた MR データを編集する事で HR かつ MR の特性を持つコンテンツにする．例として，MR データである RDF をサーバ側で編集することにより EPUB 形式で利用される XHTML と RDFa にする．XHTML+RDFa は HR であり MR の特性を持つ．
- HR 化
  著者は得られた XHTML+RDFa ファイルと CSS ファイルを保存する．必要に応じて XHTML+RDFa や CSS ファイルを改変する事が可能でなければならない．

各工程において，以上の点が重要であると考える．また、入稿，編集，出版全ての過程において，オープン性を高くし，必要に応じてシステムを変更することにより，製作の効率化と製作の幅を広げる事が可能である．

（２）実装

（１）で述べた設計に基づき実装を行う．図 4 には本アプリケーションの流れ，表 1 には要素技術について示す．

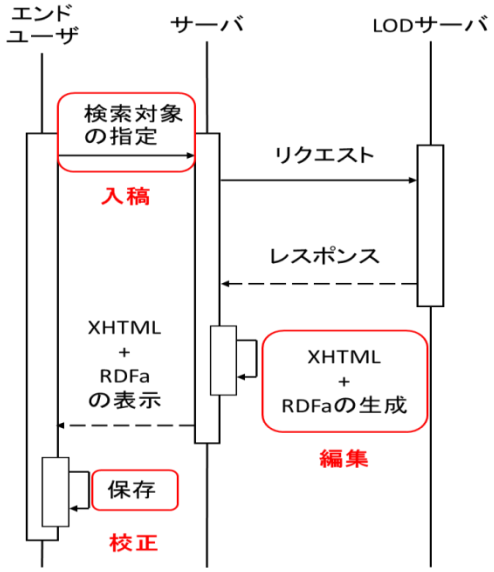

図 4.　本アプリケーションの流れ

表 1　要素技術

ユーザ側	サーバ側
・HTML5 ・XHTML+RDFa ・JavaScript ・CSS	・Java ・SPARQL ・JavaScript

a）執筆の実装

前述の設計思想に基づき，読者や編集者のニーズを柔軟かつ容易に獲得する必要がある．実装候補として，検索ボックスを設け，ユーザが入力した単語に関するオープンデータを取得する手法や，カテゴリを設けその中のコンテンツからユーザが選択をする手法が柔軟かつ容易な手法であると考える．

b）編集／校正の実装

入稿で指定した検索対象を LOD サーバに対し SPARQL(SPARQL Protocol and RDF Query Language)を使用しリクエストを行う．次に，LOD サーバ側で行われた検索のレスポンスを受け取る．この時レスポンスは RDF/XML 形式で受け取る．RDF/XML は MR なデータである．受け取った MR データを Java の XML パーサを使用し，サーバ側で XHTML+RDFa 形式に組み込む．編集されたデータは HR であり MR データである．

c）HR 化の実装

サーバ側で編集された XHTML+RDFa を CSS と JavaScript を利用し，成形した上でブラウザに表示される．読者は，得られた HR・MR データをそのままブラウザ上で閲覧する事も可能であり，ローカルに保存して必要に応じて修正を加えたり(校正)，MR データとして再利用する事が可能である．

５．オーサリング事例

4 で述べた実装に基づき，日本の城に関する図鑑を制作するオーサリングシステムの具体的な利用方法を述べる．本アプリケーションの構成は執筆までの流れは図 5，編集/校正の流れは図 6 に示す．アプリケーションの形態は Web アプリケーションとした．これは Web の持つオープン性をセキュリティの要求技術を利用して制御することで，本モデルの重要な運用が実現できるはずである．4.1 節で示したように，オープン性が高ければ，編集過程に関わる部分をすべての人が再利用する事が可能であり，製作の効率化と製作の幅が広がると言える．

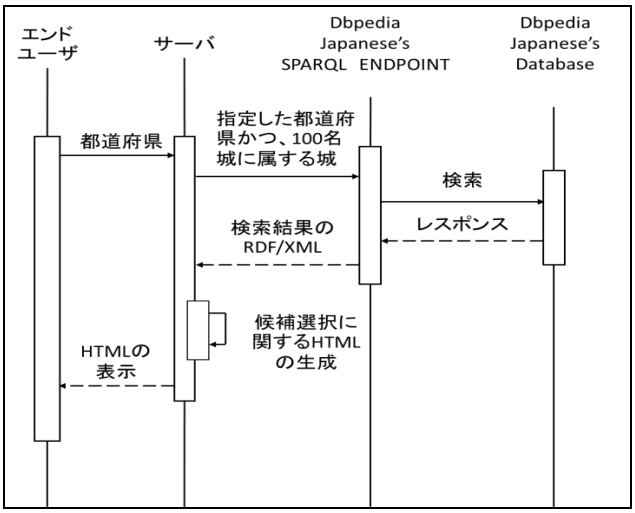

図 5.　アプリケーション構成 1

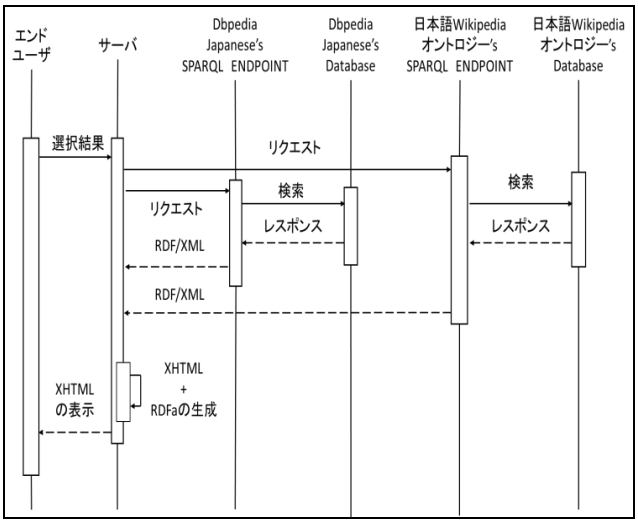

図 6.　アプリケーションの構成 2

（１）執筆

ブラウザに日本地図を表示し，ユーザに都道府県を選択させる．選択した都道府県をはっきりさせる為，JavaScript の onMouseOver 属性を利用し，選択されている都道府県の色を変化させ，視覚的に分かりやすくした．決定ボタンにより，選択した都道府県情報をサーバに送

信する．

サーバは送信された都道府県情報を元に，DBpedia Japanese の SPARQL エンドポイントに対し，「[選択した都道府県]かつ[カテゴリ:日本 100 名城]」の SPARQL を送信する．

DBpedia Japanese の SPARQL エンドポイントからのレスポンスを元に, XHTML+RDFa 化する城の候補を提案する．ユーザは表示された候補の中から，一つ選択し，サーバに送信する．

（２）編集／校正

サーバはブラウザから送信された選択結果を DBpedia Japanese と日本語 Wikipedia オントロジーの SPARQL エンドポイントに対し，選択した城の情報のリクエストを送信する．RDF/XML 形式のレスポンスをサーバが受け取ったら, XHTML+RDFa 形式で出力するための工程（編集）を行う．RDF/XML から城の図鑑として必要な，「城の名前・画像・概要・詳細」を抽出し，抽出結果をもとに，検索語の持つ繋がりを JavaScript で可視化した画像(オントロジー)を含んだコンテンツを制作する．制作後，予め用意しておいた CSS と共に，XHTML+RDFa と CSS を使ったページをブラウザ側に表示する．

（３）HR 化

ブラウザでは城を決定後，CSS・JavaScript により成形された XHTML+RDFa がブラウザ上に表示される．読者はブラウザ上で閲覧する．あるいは，必要に応じてローカルに保存し, エディターによる再編集･校正あるいは，公開による再利活用が可能である．

（４）動作例

アプリケーションの動作例を図 7,8,9 で示す．都道府県の選択から候補の選択までの流れでは，日本地図を見て都道府県を選択し，選択した県の城が表示されるという，インタラクティブ性のある選択方法である．編集過程では, 取得するデータが RDF/XML という MR なデータに対し，出力されるデータは XHTML+RDFa という HR なデータである．また，見た目は複数のページに分かれているように見えるが，実際には一つの XHTML ファイルで完結している．

（５）結果

異なるサイトで公開されている複数の Linked Open Data を活用し､インタラクティブかつ自動で電子書籍のコンテンツを執筆・編集する事が可能となった。

Linked Open Data は、関係性が明確に定義されている。データを活用し、図示する事で、視覚的に事物の繋がりだけで無く、繋がりがどのような関係を持つのか把握できる。

図 7.　入稿画面

図 8.　出力画面 1

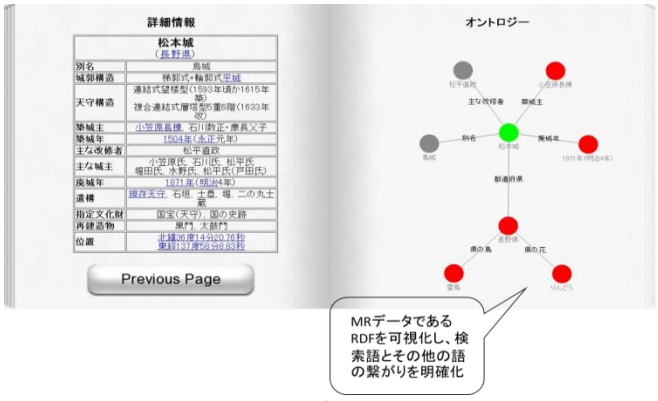

図 9.　出力画面 2

６．考察

（１）執筆

本研究では，用意されたカテゴリから入稿データを選択する手法を採用したので，著者は容易にデータを取得でき，またインタラクティブ性を持つ．しかし，本事例ではカテゴリ外のデータに関しては対応していないので，柔軟なデータの取り方を検討する必要がある．

（２）編集

MR の性質を持つ公開データを利用し，Java の技術を

利用する事で，自動的に HR/MR の特性を持つコンテンツに編集する事が可能である．また，Web アプリケーションの持つ特性により，編集の部分のオープン性に幅を持たせることが可能である．オープン性が高ければ，編集過程に関わる部分をすべての人が再利用する事が可能であり，製作の効率化に繋がる．

（３）出版

本研究では，ブラウザに備わる保存機能を使って保存する．また，再編集・校正に関しても，他アプリケーションを使わなければならない．ブラウザ上で，校正・再編集と製作過程が全て完結すれば，ユーザの負担がより軽くなると言える．

７．むすび

公開データを電子書籍に活用するという点に着目し，中でも Linked Open Data を活用した電子書籍の編集アプリケーションを提案した. 本稿では, 執筆に関して，容易でインタラクティブ性のある入稿が可能となった．編集の視点において，入稿された公開データ(MR）を自動的に HR/MR の特性を持つデータにし, LOD の特性である繋がりを意識したコンテンツを含む電子書籍が編集可能である．

読者の視点においては，LOD の持つ利点を活かし，電子書籍のページのコンテンツとして，検索語に対しての繋がりを図示化したコンテンツを得る事が出来た. また，Web アプリケーションであるため，必要なソフトがブラウザだけで良いために，利用する敷居が低いと言える．このデータを読者は再利活用可能である点は，紙での実現は難しいと言える．

今後の課題として，入稿に柔軟性を持たせる事と校正に関わる部分をブラウザ上で完結出来るようなればより，幅のある電子書籍製作に繋がると考える．

謝辞　本研究を進めるにあたり，様々なご指導を頂きました藤井章博准教授に深謝いたします．また，日常の議論を通じて多くの知識や示唆を頂いたネクストブック株式会社中山真樹様，法政大学理工学部応用情報工学科ネットワーク応用研究室の皆様に感謝します．

参考文献

1) G8 オープンデータ憲章, http://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf
2) 福井県鯖江市, http://www.city.sabae.fukui.jp/index.html
3) Linked Data - Design Issues, http://www.w3.org/DesignIssues/LinkedData.html
4) Kornschnok Dittawit, Cilas Wuwongse. A Linked Data Model for E-books, IIAI International Conference on Advanced Applied Informatics, 2012
5) 山口 琢 , 大場 みち子 , 高橋 修,電子書籍にメタデータ付きアノテーションを付けることでの想起の起点，到達点またはハブとして利用するシステムの提案，研究報告コンシューマ・デバイス＆システム（CDS）2012
6) 落合香織，三原鉄也，永森光晴，杉本重雄：マンガ Path 式を利用したソーシャル Web 上におけるディジタルマンガのアノテーション共有，第 11 回情報科学フォーラム
7) IDPF International Digital Publidhing Forum, http://idpf.org/epub
8) DBpedia Japanese, http://ja.dbpedia.org/,
9) 日本語 Wikipedia オントロジー, http://www.wikipediaontology.org/
10) Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American Magazine. Retrieved March 26, 2008.
11) LODAC project, http://lod.ac/